Appendix A

HIV Sequence Database

Search Site

Databases
Sequence DB
Resistance DB
Immunology DB
Vaccine trials DB

Publications
FAQ
Alignments
Tutorials
Reviews
Compendia
Links

Sequence DB
Search DB
Tools
HIV-BLAST
Recombination
Syn-Nonsyn
Hypermut
PCOORD
SUDI
TreeMaker
Geography
N-Glycosite
3D Structure
GeneCutter
Rip 1.9 Beta

HCV Databases

Disclaimer/Privacy

Overview of the HIV Sequence Database website

Who we are

The HIV Sequence Database focuses on five primary goals:

- Collecting HIV and SIV sequence data (since 1987)
- Curating and annotating this data, and making it available to the scientific community
- Computer analysis of HIV and related sequences
- Production of software for the analysis of (sequence) data
- Publication of the data and analyses on this site and in a yearly printed publication, the HIV sequence Compendium, which is available free of charge; subscribe here.

We have three companion databases:

- the HIV Molecular Immunology Database which provides a comprehensive, meticulously annotated listing of defined HIV epitopes
- the Vaccine Trial Database which provides a complete overview of HIV and SIV vaccine trials and their outcomes, and
- the Resistance Database, which contains a list of known drug resistance-associated mutations, their location on the genome, and associated information.

Some common questions about the contents of the database and the web interfaces are answered in the FAQ.

The database staff includes molecular biologists, sequence analysts, computer technicians, post-docs and graduate research assistants. We are part of the Theoretical Biology and Biophysics Group in the Los Alamos National Laboratory, and are funded by the Division of AIDS of the National Institute of Allergy and Infectious Diseases through an Interagency agreement with the Department of Energy. Dr. James Bradac of DAIDS/NIAID/NIH is our project officer.

What you can find on this website

The main aim of this website is to provide easy access to our database, alignments, and the tools and interfaces we have produced.

The database

The sequence database is based on HIV and SIV sequences downloaded from Genbank. We annotate these sequences with information from the literature, and sometimes from the authors. What information we add depends on what we can find, and ranges from sample information (sampling year, - country, - city), patient information (risk group, infection country and - year, sex, known epidemiological links to other patients); biological information about the virus (phenotype, tropism, coreceptor usage), technical information about the sample treatment and sequencing method, and (for a small number of important strains) extensive notes about their origin and derivation. In the future we hope to add information about treatment status of the patients and about HLA types.

At least as important as the database itself is the search interface that provides access to it. In addition to straightforward searches on many fields in the database, this tool allows the user to download alignments of certain regions, either all sequences there are for that region or a selection based on user-defined criteria. This can be very important for comparing one's sequence to existing sequences in the database; one of the most time-consuming tasks in sequence analysis used to be locating the appropriate region in sequences from the database.

Alignments

We offer meticulously edited nucleotide and amino acid alignments of all HIV-1, HIV-2/SIV, and SIV-agm genes. The new alignments of each gene contain all full-length sequences that were available for that gene at the time the alignments were made. These alignments have been pared down to contain only one sequence per person, so they are also suitable for calculating diversity in the genes and proteins. In the future we will continue to provide complete gene alignments. We make an effort to codon-align these alignments so there is an easy correspondence between the nucleotide and amino acid alignments.

The other important type of alignments we provide are the subtype reference alignments. These contain sequences that are representative for each subtype, and can be used as background alignments for trees and recombination analysis. For RIP analysis we commonly use background alignments that contain consensus sequences for each subtype, but the reference alignments can be used for this as well.

For the future we plan to provide an interface where users can automatically align their sequences to the alignments we provide.

Tools and Interfaces

We do a lot of sequence analysis in Los Alamos, and as a spin-off of this we have produced a number of programs that we think will be useful for the scientific community. For most of these a web interface is available, but some (VESPA is the most important example) have to be downloaded and run locally. This is a short list of the tools we offer:

Sequence analysis:
- HIV-BLAST runs a BLAST search against our database, which contains only HIV and SIV sequences
- TreeMaker can be used to produce simple trees; it is an interface to the Joseph Felsenstein's DnaDist, Neighbor, and Drawtree/Drawgram programs.
- SeqPublish to replace identical columns in an alignment are replaced by dashes for publication
- HXB2 Numbering Engine to find position numbers in HIV relative to HXB2
- RIP: Intersubtype Recombination Analysis, a program for detecting evidence of inter-subtype recombination.
- Search for hypermutationin a dinucleotide context
- Vespa: Signature Pattern Analysis, a program for identifying sites which are shared by one group of sequences, and are rare in another group
- SNAP: Synonymous-Nonsynonymous Analysis Program calculates syn and nonsyn values for an alignment
- Principal Coordinate Analysis (PCOORD), an interface to the program by Des Higgins for identifying patterns of correlated positions in an alignment

Other programs:
- ODprep/ODfit calculate antibody titers based on concentration and optical density data.
- HMA gel analysis, our interface to HDent and HDdist, programs for analysing data from heteroduplex mobility and tracking assays.

Questions or comments? Contact us at seq-info@t10.lanl.gov